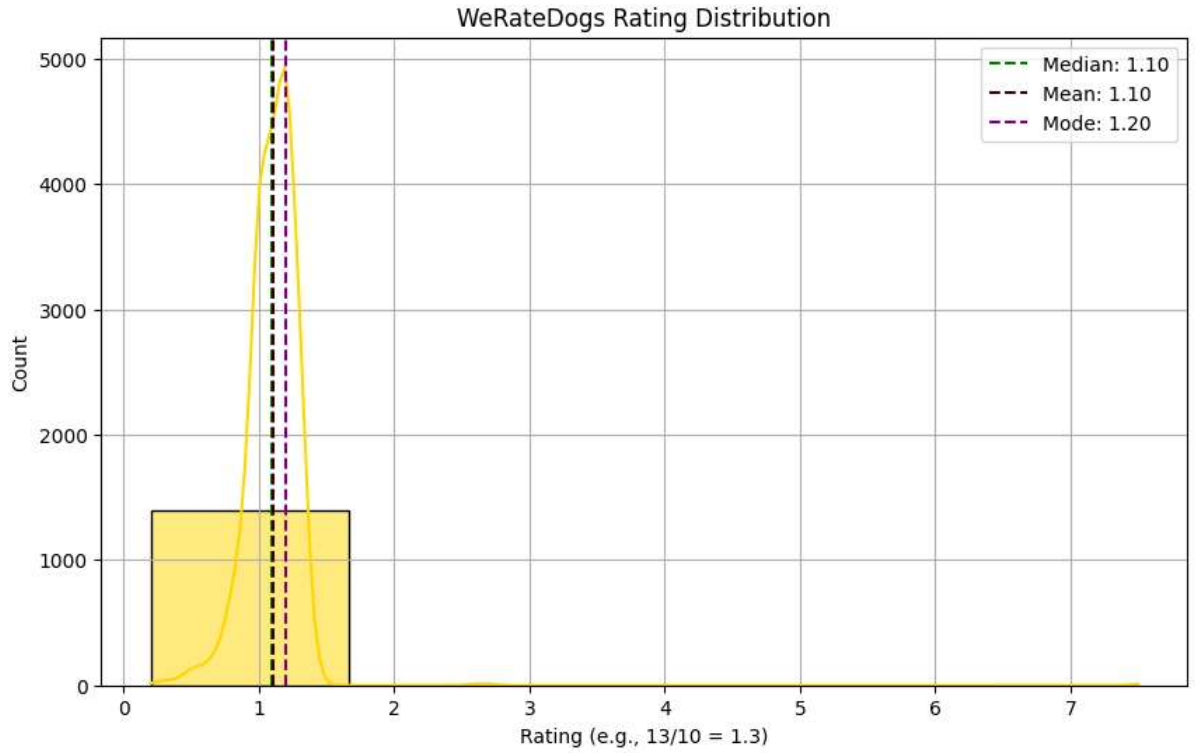
Essay on Visualisation and Insights

As part of the WeRateDogs data wrangling project, I collected, cleaned, and analysed the original Twitter archive CSV, image predictions from a neural network model, and additional tweet data pulled from the Twitter API. After the wrangling process, I moved into the analysis and visualisation phase to uncover trends and insights from the data.

One of the first things I noticed from the visualisations was how popular tweets with puppers tended to get more posts than any other dog stage. For example, pupper posts made about 136 posts, while in second place, doggo posts were only 48!—they were appreciated by the audience, who seemed to enjoy the small, cute puppers.

I also made a legend to compare the mean, mode, and median of the rating column in the master dataset. They are equal to 1.2(12/10), which is fine with WeRateDogs, which can use a higher numerator than the denominator to please watchers.



I also noticed that the number of images in each post does not affect the number of favourite posts, as two-image posts have a mean of 1300+ favourites, while the three-image posts have 1200+, and the four-image posts have 1500+ favourites!

In the end, I made a visualisation that showed golden retrievers are the most used dogs in WeRateDogs posts, while the Labrador retriever is second. As part of the WeRateDogs data wrangling project, I undertook a comprehensive process involving collecting, cleaning, and analysing various datasets. This included the original Twitter archive CSV, image predictions generated from a neural network model, and supplementary tweet data extracted using the Twitter API. Following the data wrangling phase, I progressed into a detailed analysis and visualisation stage aimed at uncovering significant trends and insights within the data.

One of the most striking observations from the visualisations was the heightened popularity of tweets featuring "puppers" compared to other dog life stages. Specifically, posts categorised as "puppers" garnered an impressive total of approximately 136 tweets, far outpacing "doggo" posts, which accumulated only 48 tweets. This data indicates a clear preference among the audience for the small, adorable puppers, suggesting that this particular category resonates more strongly with followers.

To further enhance the analytical rigour, I constructed a legend to compare key statistical measures of the rating column within the master dataset. The mean, mode, and median of the ratings all settled at a value of 1.2 (or 12/10), which aligns perfectly with the ethos of WeRateDogs. The platform’s unique approach of using a higher numerator than denominator effectively caters to its audience, providing an entertaining and engaging experience.

Additionally, I examined the relationship between the number of images in each post and the corresponding number of favourites received. Interestingly, the analysis revealed that the total number of images does not significantly influence the post's popularity regarding favourites. For instance, posts featuring two images had an average of over 1,300 favourites, while those with three received around 1,200 favourites, and four-image posts achieved approximately 1,500 favourites. This trend underscores that the charm of the content itself may outweigh the quantity of images shared.

In conclusion, my final visualisation illuminated that golden retrievers emerged as the most frequently featured breed in WeRateDogs posts, with the Labrador retriever closely following in second place. This finding highlights the popularity of these breeds among the WeRateDogs community. It opens avenues for further exploration into breed-specific trends and audience preferences within social media engagement. The insights gained from this analysis provide valuable information that can inform future content strategies and enhance audience interaction.